DOCUMENT RESUME

ED 118 633 TM 005 125

AUTHOR West, R. W.
TITLE The Summative Evaluation of Curriculum Innovations. Education Area Occasional Paper 1.
INSTITUTION Sussex Univ. (England).
PUB DATE 75
NOTE 56p.

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Course Evaluation; *Curriculum Development; *Curriculum Evaluation; Data Collection; Decision Making; *Educational Innovation; *Evaluation Methods; Instructional Materials; Models; *Summative Evaluation

ABSTRACT

After a general review of approaches to the evaluation of curriculum innovations the author presents a strategy for summative evaluation based on three related activities: (1) the intrinsic evaluation of the curriculum materials that incorporate the aims, objectives, and teaching and learning strategies of the program being evaluated; (2) a performance evaluation designed to assess the extent to which the intended outcomes of the program are achieved in action, and the level of interference from other, unintended, outcomes; and (3) a context evaluation designed to assess the effect on the curriculum proposals of the varying conditions under which they are implemented. An outline of possible techniques and methods for each of the above activities is presented, and the paper concludes by considering the whole process in relation to the types of judgement the evaluator may be required to make. Throughout the paper a strong emphasis is placed on clarifying strategic and tactical decisions when planning curriuclum evaluations, and adequate references are provided to key works of a theoretical and statistical nature. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED118633

The Summative Evaluation of Curriculum Innovations

R.W. West

TM005 125

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

R. L. WEST

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

University of Sussex Education Area Occasional Paper

1

22

THE SUMMATIVE EVALUATION OF CURRICULUM INNOVATIONS

R. W. WEST
Education Area, University of Sussex.

c 1975 R. W. WEST

Acknowledgements

The author would like to thank the following for their comments on draft forms of this paper:-

E.M. Eppel; Michael Eraut; Desmond Hogan; Jack Kerr; Colin Lacey; Janet Maw; Carolyn Miller.

Contents

1. INTRODUCTION
2. PROBLEM PRESENTATION AND ANALYSIS
3. MAJOR ISSUES IN PLANNING AN EVALUATION
   - 3.1 The characteristics of an evaluative study as distinct from a research study
   - 3.2 The goals and roles of an evaluation study
   - 3.3 The posture adopted towards the curriculum materials
   - 3.4 Goal-based or goal-free approaches
   - 3.5 Comparative versus non-comparative evaluation
   - 3.6 The evaluation of a product in a context
4. ALTERNATIVE APPROACHES TO DATA COLLECTION
   - 4.1 The 'Agricultural Botany' model
   - 4.2 The 'Social Anthropology' model
   - 4.3 The 'Interaction' model
   - 4.4 The 'Productivity' model
   - 4.5 The 'Adversary' model
5. A RECOMMENDED STRATEGY
6. TACTICAL CONSIDERATIONS
   - 6.1 Schemes for the description, analysis and intrinsic evaluation of curriculum materials
   - 6.2 The performance evaluation
   - 6.3 The context evaluation
     - 6.31 Effects on teaching and learning styles and methods
     - 6.32 Changes in the aims, objectives and attitudes of teaching staff
     - 6.33 Effects on the organisation of teaching groups
     - 6.34 Effects on the organisation of the curriculum
     - 6.35 Effects on examination and test procedures and methods
     - 6.36 Effects on the deployment, utilisation and training of staff
     - 6.37 Effects on the allocation and utilisation of resources
     - 6.38 Summary and methodological implications
7. SUMMARY AND CONCLUSIONS

Appendix 1. General Bibliography

Appendix 2. A sample curriculum profile

1. INTRODUCTION

During the last two decades we have witnessed a level of formalised curriculum development unparalleled in the history of education. In the United Kingdom the Schools' Council, the Nuffield Foundation and other funding agencies have supported curriculum development projects which by utilising a centre-to-periphery strategy have attempted to improve the general level of teaching and learning across a wide range of curriculum subjects. Whilst some of these projects have incorporated evaluation activities during the design and development stages, there has been a marked lack of investment in evaluating the effects of curriculum projects after adoption and implementation. This low level of activity is reflected in the small number of funded evaluation studies and research degree theses in curriculum evaluation or associated topics (1). The situation in the United Kingdom is further complicated by the marked lack of critical discussion and debate of the methodological problems of curriculum evaluation, and the dearth of significant publications proposing possible methodologies (2). At the same time there is evidence to suggest a degree of uncritical polarisation amongst workers in the evaluation field; a polarisation that centres on the problem of objectives and methods of data collection and analysis. This paper is an attempt to identify the key concepts and positions in curriculum evaluation, and to suggest a multi-model strategy and approach that is likely to be more

1. H.M.S.O. Scientific Research in British Universities and Colleges 1971-2, Vol. III - The Social Sciences (London, H.M.S.O. 1972); and Patterson, D. M. and Hardy, J. E. (Ed) Index to theses accepted for Higher Degrees by the Universities of Great Britain and Ireland and the C.N.A.A. Volume 20 (London, ASLIB, 1973).

2. A list of publications is presented in the Bibliography (Appendix 1).

profitable in the long term than any approach based on a single paradigm. Its main orientation is the provision of information to assist decision makers concerned with the general process of curricular improvement, rather than those making adoption, adaption or rejection decisions within a specific school. The particular strategy suggested has been tested in an evaluation of one of the Nuffield 'O' Level Science projects (3), but any references to that study are made for illustrative purposes only.

2. PROBLEM PRESENTATION AND ANALYSIS

The existence of a wide variety of curriculum proposals immediately raises a range of questions for teachers, adminstrators and parents, in their roles as providers and consumers within the educational system (4). Firstly, the potential consumer of a curriculum package is concerned with questions of ends, of the goals of the proposals, and their compatability, or incompatability, with his own educational goals. He requires to know, in other words, the precise intentions of the proposals. Secondly, he is likely to be concerned with problems of means, the methods and strategies whereby the goals of the curriculum proposals are to be achieved. These 'questions of means' are in part philosophical/pedagogical questions about the assumptions made regarding content and teaching and learning strategies, and in part organisational/economic

---

3. West, R. W. An Evaluation of the Nuffield Science Teaching Project Ordinary Level Chemistry Proposals: Text, Performance, and Context (University of Sussex, unpublished D. Phil. thesis, 1974).

4. It is convenient to regard parents, rather than pupils, as the prime consumers at this stage, although this does not pre-empt the acceptability of the curriculum proposals to pupils as being a major evaluative question.

questions relating to the deployment and utilisation of educational resources such as staff, facilities, equipment, and space. Thirdly, he will be concerned with the status of the curriculum proposals, especially with respect to the amount and quality of developmental work that has been undertaken prior to publication. Finally, the potential consumer is interested in questions of effect; the possible effects adoption will have on his own practices and, of greater long term importance, the effect adoption will have on the educational progress and development of his pupils, or his children. The relationship between these multiple questions of end, means, status and effect is dynamic and forms the ground for the creation of an evaluative scheme.

3. MAJOR ISSUES IN PLANNING AN EVALUATION

3.1. The characteristics of an evaluative study as distinct from a research study

Many research studies in the context of education are essentially examples of the hypothetico-deductive method of inquiry applied to clearly defined problems with a view to determining the simplest, most parsimonious, explanation of events with the highest level of generalisability. The attainment of these ends requires a strong theoretical base, the clear specification of hypotheses, precise definition of test instruments and other data gathering devices, and the careful control of the experimental environment and conditions. Implicit in this process is the careful control of the value judgements of the researcher which are usually restricted to the selection of the problem. The activity is paradigmatic in Kuhn's sense (5) and rationalist in Stake's sense (6).

5. Kuhn, T. S. The Structure of Scientific Revolutions (Chicago, University of Chicago Press, 2nd ed. 1970).

6. Stake, R. E. 'Language, rationality and assessment' in Beatty, W. H. (Ed) Improving Educational Assessment and An Inventory of Measures of Affective Behaviour (Washington, Association for Supervision and Curriculum Development, N.E.A., 1969) pp. 26-28.

An attempt to study an educational programme in action conflicts with many of the assumptions underlying the above notions of a research study. Firstly, the problem is almost totally constrained by the situational context of the programme and the inability to control or manipulate that context. Secondly, the lack of a strong theoretical base for the study, and often for the programme itself, leads to conflicting views of appropriate methodology, techniques and instruments, and the requirement to make value judgements explicit in the definition of both problems and methods. Thirdly, the study is almost certainly undertaken for overtly utilitarian rather than intrinsic reasons. Fourthly, the findings of a study which is programme and context specific are usually generalisable only within very narrow limits. So instead of leading to a parsimonious explanation of reality, the study will probably err towards complexity and detail, even at the expense of some redundancy. Finally, the study will be pre-paradigmatic, located within Kuhn's notions of pre-science and natural history (7) and rely heavily on a heuristic and empiricist approach to problem definition, observation and interpretation. An evaluation contains many of these elements if one accepts Stake's broad working definition that

> As evaluators we should make a record of all the following: what the author or teacher or school board intends to do, what is provided by way of an environment, the transactions between teacher and learner, the student progress, the side effects, and last and most important, the merit and shortcoming seen by persons from divergent viewpoints (8).

---

7. Kuhn, T. S. op. cit.

8. Stake, R. E. op. cit. p. 15.

In other words an evaluation is characterised by the range of its objectives, the variety of data collection devices employed, the scope of the data collected, and the characteristics of the conceptual framework imposed on that data for the purpose of description, inference and generalisation. The overall level of generalisability is likely to be limited but should be directed to the improvement of educational decision making.

3.2 The goals and roles of an evaluation study

The development of an appropriate scheme to achieve the elements of an evaluation as defined in the previous section requires a careful analysis of the distinction between the goals of the evaluation study and its role or roles. Scriven (9) differentiates between these two functions by suggesting that goals are essentially common across all evaluative studies, being related to questions of merit, worth and value, e.g. 'How well does a programme perform?; does it perform better than another programme?; is the programme worth the money it costs?' The role of evaluation is however context dependent and is primarily related to the potential utilisation of the product of the evaluation activity. Thus it may be directed towards identifying areas for possible improvement in a teaching and learning sequence or scheme; or towards identifying basic problems of acceptance and implementation; or the clarification of outcomes and the consequent build up of continuation studies. A major distinction in roles introduced by Scriven is that between formative and summative evaluation and the interrelationship between them. Whilst Cronbach

9. Scriven, M. 'The methodology of evaluation' in Tyler, R. W., Gagne, R. M. and Scriven, M. Perspectives of Curriculum Evaluation (Chicago, Rand McNally, 1967) pp. 40-43.

and others (10) have tended to over-emphasise and value formative evaluation, evaluation contributing to the development of a course or programme before completion. Scriven suggests that summative, or outcome, evaluation is important in providing potential consumers with evidence of worth prior to their making adoptive, or rejective, decisions. In the context of many studies, the basic role adopted is likely to be summative as the curriculum proposal will often have been published prior to the commencement of the evaluation. Within this basic position a formative perspective will exist in that many curriculum proposals are seen by their authors as a contribution to a continuing process of curriculum reform.

3.3 The posture adopted towards the curriculum materials

A further consideration in developing an appropriate methodology is that of the position adopted in relation to the curriculum materials themselves. Firstly, curriculum materials can be evaluated intrinsically, a process which centres on the general appraisal of the curriculum itself by means of an analysis of its content, goals, teaching and learning strategies, and methods of assessment. The basic questions in an intrinsic or 'armchair' evaluation are those of coherence and internal consistency and these are not evaluated empirically. Whilst a comprehensive intrinsic evaluation of a curriculum project presents problems both through the introduction of

10. See for example, Cronbach, L. J. 'Course improvement through evaluation', Teachers' College Record, 64 (1963), pp. 672-683; and Flanagan, J. C. 'The uses of education evaluation in the development of programs, courses, instructional materials and equipment, instructional and learning procedures, and administrative arrangements' in Tyler R. W. (Ed) Educational Evaluation: New Roles, New Means (Chicago, University of Chicago Press, 1969).

intermediate, or evaluator created goals, and from the standpoint of the development of appropriate analytical schemes, it is clear that unless such an analysis is incorporated in the summative evaluation of a project many effects and value positions will be overlooked. This is particularly important when evaluating materials which do not incorporate a comprehensive statement of explicit aims and objectives.

Secondly, curriculum materials can be evaluated in 'pay-off' (11) or performance (12) terms, the traditional measurement or estimation of the effects of the materials on the learner. Unlike intrinsic evaluation, in which criteria are not usually operationally formulated, the procedures for a performance evaluation aim at estimating differences in performance on prespecified criteria. The synthesis between intrinsic and performance evaluation is generated by the establishment of the intended and/or likely outcomes of the projects and their relationship to the actual outcomes established in the field. An intrinsic evaluation is therefore hypothesis forming and a performance evaluation hypothesis testing. The first is essential not only to the generation of hypotheses, but to the evaluation of their congruence within the general structure of the curriculum itself.

11. Scriven, M. op. cit. p.54.

12. Kraut, M. R. 'The role of evaluation' in Taylor, G. (Ed) The Teacher as Manager (London, N.C.E.T., Councils and Education Press, 1970) pp. 114-128.

3.4 Goal-based or goal-free approaches

In discussing the derivation of suitable hypotheses to test in a performance evaluation it should be noted that controversy exists between proponents of goal-based evaluation techniques and the more recent suggestion that evaluation should be goal-free (13). Central to the concept of intrinsic evaluation is the examination of the curriculum materials in order to identify the intended goals of the producer, and to evaluate the likelihood that these goals will be achieved by implementing the course according to the procedures detailed by the producer. In addition to the goals made explicit by the curriculum proposers it should be noted that the evaluator may introduce his own goals into the evaluative scheme. These may be derived from the intrinsic evaluation of the course materials as alternatives to the goals suggested by the proposers, or be based on assumptions made regarding the probable effects of the proposals in the teaching and learning situation. These assumptions may be made either on the basis of the evaluator's own experience as a teacher, or on discussions with other teachers of the subject. The goals, however they are derived, will serve as the basis for the development of the test instruments or questionnaires used in any performance evaluation. Thus, according to the model, if the curriculum intends to 'enable the pupil to use a chemical balance with a high degree of accuracy', the performance evaluation scheme would include a test of ability to use a chemical balance. Critics of the behavioural objectives approach to the definition of educational outcomes have suggested that the clear definition of the intended outcomes of

13. Scriven, M. 'Prose and cons about goal-free evaluation' Evaluation Comment, Vol. 3, No. 4, (1972) pp. 1-4.

curricula is no guarantee that these outcomes will be attained, and that there is little evidence to support the view that teachers teach better when objectives have been stated clearly (14). Scriven, in placing heavy emphasis on the role of intrinsic evaluation, stresses the requirement to estimate the secondary, unintended, or side-effects of a curriculum, again in terms of goal statements, and to ensure that the level of attainment of these is checked at the 'pay-off' stage of the evaluation. It is significant that the problem of implied value judgements between intended and unintended effects has led Scriven to propose goal-free evaluation procedures within the context of both formative and summative evaluation (15). In a goal-free approach test instruments and questionnaires used in a performance evaluation would reflect a full range of reasonable outcomes of a particular curriculum project, as distinct from the more limited sub-set of outcomes planned by the project team. Scriven argues strongly that goal-free evaluation safeguards against tunnel-vision effects implicit in goal-based evaluation; the over-preoccupation of the evaluator with estimating the achievement of stated goals, at the risk of failing to observe strong unintended effects which may be more educationally desirable than the intended

14. See, for example, Atkin, J. M. 'Some evaluation problems in a a course content improvement project' Journal of Research in Science Teaching, Vol. 1 (1963), pp. 129-132; and Popham, W. J. 'Performance tests of teaching proficiency: rationale, development and validation'. American Educational Research Journal, Vol. 8 No. 1 (January 1971) pp. 105-117.

15. Scriven, M. Evaluation Comment op. cit. p. 1.

ones (16). In developing a strategy for the performance evaluation of Nuffield 'O' level Chemistry the present writer noted the goals of the Nuffield Chemistry Project team, but introduced additional goals based on his experience of teaching a Nuffield based chemistry course. In addition a survey of pupil attitudes to science was incorporated in the performance evaluation as an essentially goal-free component in that it reflected a classification of attitudes developed by the National Foundation for Educational Research (17), rather than any statement of affective objectives made by the Nuffield project team.

3.5 Comparative versus non-comparative evaluation

The evaluator of a curriculum programme has the option of examining that curriculum in isolation from any other approach to teaching and learning in that field, or can set out from the start to evaluate the programme by comparison against an existing programme (usually the 'new' against the 'old' or 'traditional'). A major proponent of single group evaluation studies is Cronbach who has stated his position as

> Since group comparisons give equivocal results, I believe that a formal study should be designed primarily to determine the post-course performance of a well defined group, with respect to many important objectives and side effects (18).

16. Note should be taken however of the strong opposition to goal-free approaches. See, for example, Stufflebeam, D. L. 'Should or can evaluation be goal-free?' Evaluation Comment, Vol. 3, No. 4 (1972) pp. 4-5.

17. National Foundation for Educational Research; Pupil Opinion Poll: Science No. 104 (Slough, N.F.E.R., 1968).

18. Cronbach, L. J. op. cit. p. 238.

Cronbach argues that results showing that programme A is superior to programme B do not necessarily give insight into the reasons for superiority, reasons which would be more apparent from a non-comparative study utilising a large pool of test items. He also suggests that studies based on gross comparisons on a limited range of variables can only show small differences. Scriven argues however that the failure to identify the causes of differences between programmes is a function not of comparative studies, but of a concentration on performance evaluation at the expense of a commitment to prior intrinsic evaluation; and that the failure to show significant differences between groups is not, a priori, a failure to identify crucial information about the performance of a project providing good tests of the criterion variables are used. 'No difference' is not 'no knowledge'. Whilst supporting Cronbach's view that evaluation must involve more and more detailed micro-studies incorporating the evaluation of a range of criterial parameters, Scriven argues that this can be undertaken in the context of comparative studies. Scriven furthermore sees the element of comparison being implicit in every evaluation and concludes that comparative evaluations are

> often very much easier than non-comparative evaluations, because we can often use tests which yield differences instead of having to find an absolute scale and then eventually compare the absolute scores (19).

19. Scriven, M. 'The Methodology of Evaluation' op. cit. p. 64.

3.6 The evaluation of a product in a context

Whilst the intrinsic evaluation of a curriculum project involves a priori judgements related to the evaluator's knowledge and experience of the likely context in which the product and a given educational institution or environment. Nevertheless these factors are important and a full evaluation strategy would involve analysis of the effects on the things that surround the learner particularly within the context of the performance evaluation. As Taba suggests

> Failure to assess realistically the effect of the existing conditions has often led to the discrediting of a given curriculum design when the difficulty may not have been in the design but in the discreptancy between the requirements of the design and the conditions for implementing it (20).

Thus we are not simply concerned with the merits or demerits of the curriculum project and estimating the gap between what it sets out to do and what it achieves, but have also to consider the demands made on other pupils, teachers, ancillary staff, resource allocations and the institutions of education at large (21). These might include the identification of conflicts in the

20. Taba, H. Curriculum Development: Theory and Practice (New York) Harcourt, Brace and World, 1962) p.426.

21. Lortie argues that project evaluators should be concerned not only with the immediate effects of a project on a school but also its congruence with the educational value system within which the school is located. He argues therefore for the broader definition of goals and their alignment with specific social goals. See Lortie, D. C. 'The cracked cake of educational custom and emerging issues in evaluation' in Wittrock, M. C. and Wiley, D. E. The Evaluation of Instruction: Issues and Problems (New York, Holt, Rinehart and Winston, 1970), pp. 149-164.

deployment of staff and resources resulting from the introduction of a new course; incompatibilities between the teaching and learning methods of the new course and those of complementary and other courses; and problems relating to the organisation, structure and objectives of courses that preceed, or follow on from, the proposed new course. In addition note should be taken of the relationship between the new course and the expectations of parents, examiners, employers and other areas of education. Making evaluative judgements on the interactive effects between the 'achievements of the course' and the 'effects on the environment' could involve forms of cost-benefit or cost-effectiveness analysis, and may lead to either changes in the project or the context. It is however possible to consider the use of more descriptive analyses of effects on the environment by introducing questions relating to changes in curriculum structure, staff work loads, and resource allocations subsequent to the adoption of the curriculum project. Alternatively, on the assumption that similar secondary schools would normally be expected to be exposed to similar staff, curricular and resource constraints, differences in patterns of resource allocation in schools adopting the new curricula and others rejecting the innovation may be causally related to the demands of the innovation.

4. ALTERNATIVE APPROACHES TO DATA COLLECTION

As indicated in the previous section the performance evaluation component of an evaluation will be concerned both with the extent to which the course attains its objectives, and the effects the course has within a defined educational context. This bifurcation of aims requires a careful consideration of the approaches to data collection. Five distinctive models are discussed briefly in this section.

4.1 The 'Agricultural Botany' model

The traditional model for course evaluation is firmly grounded in the psychometric models developed for the testing/evaluation of individual pupils/students, mainly for purposes of diagnosis, selection or guidance; and the experimental or quasi-experimental designs for experiments in psychology and education based on the statistical work of Fisher and others (22). The only major variation in mehodology introduced during the transfer from testing and research contexts into the evaluation context has been the movement away from norm-referenced tests and the development of criterion-referenced tests. Parlett, in summarising the nature of the 'Agricultural Botany Paradigm' which he equates with 'evaluation as testing' states that

> This description of evaluation procedures would not be complete without first specifying what would be its most supreme creation, its ne plus ultra. It would take the following form: large, finely balanced samples are formed into experimental and control groups; they are tested before the pedagogical 'treatment' is applied; and tested again afterwards; 'before and after' and 'between samples' comparisons can then be drawn (23).

Criticism of the above model is directed at the problems, and even desirability, of defining objectives as the essential pre-cursor to the development of the test instruments; and the effects the constraints of experimental design may have in determining the range of variables considered and the context in which they are examined. Put at its worst the evaluator, in order to safeguard the validity of

22. Summarised effectively in Lindquist, E. F. (Ed) Educational Measurement (Washington, D. C. American Council in Education, 1951).

23. Parlett, M. 'Evaluating innovations in teaching' in Butcher, H. J. and Rudd, E. (Ed) Contemporary Problems in Higher Education (London, McGaw-Hill 1972) p. 146.

his design, may be forced to carefully control the independent variables and limit the range of dependent variables to functions or characteristics which can be measured with high validity and reliability. That the resulting experimental situation, with its careful control of 'interference and noise', may be unrealistic in terms of real children and schools has led Parlett to describe it as 'A paradigm for plants, not people' (24). A final category of criticism of the traditional psychometric paradigm centres round the extent to which a tight, rational paradigm can be used when the level of knowledge as to the relationship between ends and means is so incomplete. Where knowledge is loosely framed in Bernstein's sense, (25) or pre-scientific in Kuhn's (26), true experimental designs may be considered inappropriate and broader techniques such as survey analysis or panel analysis may be more appropriate.

4.2 The 'Social Anthropology' model

An alternative model for curriculum evaluation is based on the notions of evaluation as interpretation (27) or illumination (28) both of which have roots in a social anthropology paradigm which is essentially inductive, interpretive and based on related a posteriori

24. Parlett, M. op. cit. p. 146.

25. Bernstein, B. 'On the classification and framing of educational knowledge' in Young, M. F. D. (Ed) Knowledge and Control (London, Collier-Macmillan, 1971) pp. 47-69.

26. Kuhn, T. S. op. cit.

27. Parlett, op. cit. pp. 150-153.

28. Trow, M. 'Methodological problems in the evaluation of innovation' in Wittrock, M. C. and Wiley, D. E. op. cit. pp. 289-305.

judgements. In this context the evaluation would be concerned with a close, in-depth study of a curriculum in action in a limited context, say one school, with the intention of eliciting as wide a data base as possible from pupils, teachers, parents, employers, administrators and the curriculum developers themselves. Techniques employed might range from interviews, conversations, questionnaires, check-lists, observation schedules an' the analysis of children's written work. The evaluator would seek to create a coherent picture of the outcomes of the innovation in terms of effects on a highly complex system of relationships. The pre-specified design for the study would consist essentially of broadly framed research questions and some preliminary planning schedules for the conduct of the study. Proponents of the 'social anthropology paradigm' argue for its closer relevance to the real issues in curriculum development and innovation which are largely related to establishing the way the intentions of the developer are interpreted, and thus modified, by the teacher; and to the ability of the evaluator to readily observe unplanned side effects or secondary outcomes. Providing the participant observation is undertaken with tact and care, the level of distortion of the system by the act of observing it is low, certainly much lower than the distortion introduced by creating true experimental situations.

Critics of the paradigm identify the dangers of an excess of irrelevant qualitative data preventing the evaluator from establishing relevant links and relationships; the fallacy of assuming that the non-specification of learning outcomes makes it any easier to spot unintended outcomes and the problem of effectively quantifying much

of the anecdotal data collected (29).

4.3 The 'Interaction' model

The problem of evaluating the effects of a given teaching and learning sequence, project or programme can be approached through the detailed observation and analysis of the transactions that occur in the process of teaching and learning itself. In other words the nature of the interaction between teacher and learner, learner and learner, or learner and resources can be categorised, quantified, analysed and interpreted in order to create a model of relationships between the curriculum materials, methods, and the educational context. Thus the teacher, pupils and the curriculum are seen as an interactive system. Two major problems exist in the design and development of an evaluative scheme based on interaction analysis. Firstly there is the question of the nature of the interactions that are to provide the observational data. Secondly there is the problem of the criteria for establishing the classificatory scheme. In terms of the nature of the interactions the majority of approaches have been based on the analysis of the verbal intercourse of a lesson, or series of lessons, and utilise concepts developed initially by Flanders and Bellack (30). Thus the language of question and response, and of general classroom talk, provides the data base for the study, with an underlying theoretical assumption

29. For a general discussion of methodolgoy and the problems see Becker, H. S. Problems of Inference and Proof in Participant Observation. (London, Allen Lane and Penguin Press, 1971) pp. 25-38.

30. Flanders, N. A. Teacher Influence, Pupil Attitudes and Achievement (Co-operative Research Monograph No. 12, U.S. Department of Health, Education and Welfare, 1965); Flanders, N. A. Analysing Teaching Behaviour (New York, Addison-Wesley, 1970); and Bellack, A. A., Kliebard, H. M., Hyman, R.T. and Smith, F. L. The Language of the Classroom (New York, Teachers' College Press, 1966).

that the language forms used by the pupils are related to the quality of their understanding of concepts and techniques, or their affective attitudes towards the curriculum materials.

Approaches to the classifactory problems have either been inductive in that categories have been created ab initio from data comparable to that which will be classifed by the developed system; or deductive in that a priori categories are derived from an appropriate process or cognitive model and subsequently developed against the cutting edge of appropriate data. In the case of the Earth Science Curricular Observation Instrument the process of developing the observation schedule involved the inductive derivation of probable objectives for the curriculum itself, and thence to the deductive specification of appropriate behavioural categories and the development of the schedule (31).

Methods of observation utilised in interaction studies have ranged from participant observation and the completion of checklists, to the post event analysis of audio-tapes or video-recordings of the total transactions. Some schedules include provision for non-verbal interactions such as facial expressions, but no schedule, to the writer's knowledge, incorporates interactions with resources such as experimental equipment, book resources, audio-visual resources or models.

The major methodological problems with interaction models for curriculum evaluation appear very similar to those discussed under

---

31. Smith, J. P. 'The development of a classroom observation instrument relevant to the Earth Science Curriculum Project', Journal of Research in Science Teaching, Vol. 8, No. 3 (1970) pp. 231-235.

previous headings; namely validity, reliability and interference are effectively summarised by Gallagher, Nuthall and Rosenshine (32) and by Galton, Eggleston and Jones (33).

4.4 The 'Productivity' model

If the adoption and implementation of a set of curriculum proposals is conceptualised in input-output terms, it may be appropriate to base an evaluation within an economic framework and thus estimate the benefits, or outcomes, of the course in terms of the costs of inputs. At the macro level of curriculum decision making consideration of costs and their relationship to the nature and quality of outputs has led to general notions of accountability, and particular systems such as performance contracting where unit costs related to specified levels of pupil performance are established in advance of the adoption of a teaching and learning system. The theoretical basis for attempts to analyse curricular factors in productivity terms is formed in the economic models of cost-benefit and cost-effectiveness analysis. Within this set of analytical models Bowman (34) has suggested three distinctive forms of analysis:-

32. Gallagher, J. J., Nuthall, G. A., Rosenshine, B. Classroom Observation (Chicago, Rand McNally, 1970).

33. Eggleston, J. F., Glaton, M., Jones, M., Contexts of Science Learning (Schools' Council Project for the Evaluation of Science Teaching Methods, University of Leicester, unpublished). An advance copy of the draft submitted to the Schools' Council has been extremely valuable in preparing this paper.

34. Bowman, M. J. private memorandum quoted in Thomas, J. A. The Productive School, A Systems Analysis Approach to Educational Administration (New York, Wiley, 1971) p. 82.

a) Simple cost-benefit analysis, in which the outputs of the system the benefits, can be priced and the unit costs of unit benefits established. The benefits of an educational activity may, for example, be expressed in terms of levels of employment with incomes amortised over a full working career taken as indices of benefit. The relationship between input costs and output benefits is therefore expressed in terms of unit monetary costs by unit monetary benefit.

b) Simple cost-effectiveness analysis, in which inputs can be costs but the outputs of the system cannot be priced although they can be described in terms of defined goals or objectives, either singly or as a weighted index based on a combination of objectives. Thus the analysis may consist of a model for estimating the unit cost of the attainment of a fixed benefit, or for estimating the variable benefits achievable for a fixed cost. This model is probably most appropriate when considering a range of alternative teaching strategies for attaining a given set of agreed educational objectives. For example, one may wish to determine the most effective method of teaching elementary data handling techniques and therefore compare a traditional class teaching method, a method based on a self-instructional learning programme, and an audio-visual tutorial method. The costs of implementation of each method would be estimated and that attaining the learning objectives at an agreed level for the lowest cost would be the method adopted.

c) Complex cost-benefit or cost-effectiveness analysis, in which the unit costs are estimated for benefits that are realised in varying

degrees and which may, or may not, be expressed in monetary terms.

All three models therefore depend on an ability to effectively estimate the input costs of teaching and learning, with the major distinction between cost-benefit and cost-effectiveness analysis being the ability to realistically translate, or equate, outputs into monetary terms. Cost-benefit analysis assumes that such a translation can be made.

The problems of developing a productivity model for curriculum evaluation centre therefore on ways and means of describing clearly the outputs of the curriculum under examination, either in terms of the levels to which objectives are attained, or in terms of benefits such as continuing involvement in education, levels of employment, or estimates of job-satisfaction or efficiency. In a comparative evaluation, where one curriculum or course is compared with another, care must also be taken regarding the extent to which any unitary set of objectives across which the two courses are to be compared are legitimate and expected outcomes of both courses. Where different objectives are attained the problem of making value judgements between objectives would create major difficulties particularly if a system of monetary equivalency was envisaged.

4.5 The 'Adversary' model

The clue to the notion of an 'adversary model' lies in Stake's view that evaluators should record 'the merits and shortcomings (of the project) seen by persons from divergent viewpoints' (35).

35. Stake, R. E. op. cit. p. 15.

Levine and Kourilsky (36) mutually explore the possibility of developing an approach to evaluation based on an analysis of the curriculum project from an overtly supportive point of view and an overtly negative one. It is argued that current practice centres on the 'single recommendation approach' and that this is acceptable when:-

a) the client and evaluator have a strong prior agreement about ends but doubts over means;
b) there is basic agreement over the interpretation of the data on which the evaluation recommendation is made;
c) the decision maker (client) leaves problems of validity to the evaluator;
d) the recommendations of the report are being used for accountability decisions rather than future large scale policy (spending) decisions (37).

Kourilsky argues that where these conditions are not met, and particularly where large scale commitment to resource spending is anticipated, the one view of the product can create major problems. In the adversary model the presentation of the two views allows the decision maker to benefit from

36. Levine, M. 'Scientific method and the adversary model: some preliminary suggestions' and Kourilsky, M. 'An adversary model for educational evaluation', Evaluation Comment, Vol. 4 No. 2 (1973) pp. 1-6.

37. The Levine and Kourilsky paper assumes, as does much of the American evaluation literature, that a client-professional evaluator system exists, in which the client is a school board director or some other consumer of an educational product, i.e. purchaser.

a) the explicit judgemental debate, or legal argument, between the 'Affirmative Evaluator' and the 'Negative Evaluator' each of whom are presenting prima facie cases for their categories of appointment; and

b) the dialetic of the debate which is likely to produce a synthesis of view. It should be stressed in describing this model that only one programme is presented and the debate does not centre round a programme and counter-programme, e.g. BSCCS or Nuffield Biology, Nuffield Biology or 'Traditionalist Biology'.

Whilst the Kafka-esk nature of the debate might create problems it is clear that this model is of limited application in the United Kingdom where the role of summative evaluation as an aid to education decision-making by resource allocators (e.g. Chief Education Officers, Spending Officers or Head-Teachers) has yet to be demonstrated.

5. A RECOMMENDED STRATEGY

On the assumption that the overall aim of an evaluation study is to provide teachers, administrators and curriculum developers with evidence on which decisions relating to the adoption, adaptation or further development of a curriculum can be made, the following strategic decisions can be taken in the light of the above analysis.

A full summative evaluation is conceptualised as a process which incorporates:-

a) an intrinsic evaluation of the published curriculum materials designed to identify and evaluate:-

   i) both the intended and unintended outcomes of the course;

   ii) the main curricular assumptions of the authors;

   iii) the proposed teaching and learning strategies;

iv) the major resource and organisational implications of the course.

b) a performance evaluation designed to assess the extent to which the intended outcomes of the course are achieved, and the level of interference from unintended outcomes. This evaluation would ideally be comparative, in that the performance of pupils following both the new and existing courses would be considered, and be based on the 'agricultural - botany' model described in Section 4.1 although a tight control of interacting variables would not be attempted.

c) a context evaluation designed to assess the effects of implementation of the course in terms of staff loads, resource allocations and curricula organisation in schools; and estimate the effect on the curriculum proposals of the varying conditions under which it has been implemented. The context evaluation would thus contain elements of both the 'social anthropology' and 'productivity' models described in Sections 4.2 and 4.4 respectively.

6. TACTICAL CONSIDERATIONS

The implementation of the above tripartite evaluation strategy raises a number of tactical problems which are appropriately considered at this stage. These relate to the development of appropriate techniques for the intrinsic evaluation of the curriculum materials, the development of an appropriate methodology for the performance evaluation, and the analysis of the contextual effects of the curriculum in action.

6.1 Schemes for the description, analysis and intrinsic evaluation of curriculum materials

The intrinsic evaluation of curriculum materials involves a detailed analysis of publications and associated materials with a view to identifying and evaluating their most important characteristics and implications for pupils, teachers and schools. The analysis is essentially concerned with establishing the aims and objectives of the curriculum package, and judging the extent to which it is likely to attain those objectives and at what cost, expressed in terms of effects on the school, its curriculum and its resources. In additon by concentrating attention on ascertaining both the intended and unintended outcomes of the curriculum package, the analysis would enable the objectives of both the performance and context evaluation to be precisely defined. In other words the decisions regarding extrinsic evaluation are dependent on the results of the intrinsic evaluation.

In a general review of schemes for the analysis of curriculum materials Eraut, Goad and Smith (38) have identified three specific functions for an analysis. These are:-

a) A Descriptive - Analytic Function directed at describing materials and elucidating their rationale and structure;
b) An Evaluative Function in which materials are judged against a range of criteria irrespective of any specific context in which they may be used; and

38. Eraut, M. R., Goad, L. H., Smith, G. E. The Analysis of Curriculum Materials (Sussex, University of Sussex Education Area Occasional Paper No. 2, 1975) p. 23.

c) A Decision Making Function which is context specific and advisory in terms of selection or implementation decisions.

Clearly in a fully integrated scheme of summative evaluation all three functions are important.

Two schemes of analysis to emerge from the curriculum development movement in the United States which are relevant to this discussion are those published by the Social Science Education Consortium (39) and the Far West Laboratory for Educational Research and Development (40). Both schemes are essentially concerned with the descriptive-analytic function only and consist of a structured sequence of questions directed at describing and analysing the rationale and implications of curriculum materials. The S.S.E.C. scheme was originally designed for use in conjunctions with the Consortium's information bank and services, and was directed towards curricula in the social sciences at secondary level. The Far West Laboratory scheme was specifically desgined for the analysis of elementary science curricula but, in common with the S.S.E.C. scheme, can be applied to any subject or age group. It should be emphasised that both schemes are purely descriptive and contain no provision for the quantification of data, except in relation to the use of time and the costs of publications, materials or equipment. Furthermore although the Far West Laboratory scheme is neutral in its structure, the S.S.E.C. scheme is firmly rooted in the Tyler-Bloom model for curriculum

39. Stevens, W., Morrissett, I. 'A system for analysing social science curricula', EPIE Forum, Vol. 1 No. 4 (December 1967) pp. 10-15.

40. Hutchins, C. C. (Ed) Science - a Process Approach: Programme Report (Berkeley, Far West Laboratory for Educational Research and Development, 1970).

development and requires that statements of curriculum objectives be made in behavioural form. A major disadvantage of both schemes however is the absence of a specifically evaluative section.

The approach to intrinsic evaluation adopted by the Volkswagen Curriculum Analysis Project (41) attempts to overcome these problems by combining a description of the curriculum materials with an analysis of the strategic implications of the 'curriculum in action' and an evaluation of both the materials and the strategies. The strategic section, which is based on a modified Tyler model, is primarily concerned with identifying the range of key curriculum decisions that will be involved when adopting the materials in typical, but generalised school contexts. These decisions are grouped under five headings referring to aims; subject matter; objectives and outcomes; teaching, learning and communication strategy; and assessment pattern. Much of this analysis centres on comparing typical school aims, objectives and methods against those proposed by the authors of the curriculum materials under review and the scheme, as presented, has a strong goal-free perspective. The approach to the evaluative function is similarly intrinsic in that it concentrates on the analysis of information from project trials, tests, evaluations and reviews. However, the full Volkswagen scheme is extremely lengthy and complex and whilts its effective application produces a full and illuminating intrinsic assessment of the likely outcome of adopting a given curriculum, a combined intrinsic, performance, and context evaluation has additional objectives.

41. Eraut, M. R., Goad, L. H., Smith, G. E., op. cit.

The following analytical scheme which was developed from the 1967 S.S.E.C. report attempts to incorporate key aspects of both the descriptive - analytical and evaluative functions, but does so with a full performance and context evaluation in mind. The main differences in aim between this and the Volkswagen scheme are

a) the absence of any criticism of the rationale and strategy of the curriculum other than that implied by internal inconsistency; and
b) the need to use the intrinsic evaluation to inform the evaluator rather than a broader audience.

The scheme consists of five related sections, namely:-

Section 1 A description of course materials.

Section 2 An anlysis of the antecedent conditions implied by the course materials and/or the course proposers. Sub-divisions within this section deal with assumptions about pupils, teachers and the curriculum and organisational context within which the course is expected to operate.

Section 3 An analysis of the rationale and strategy assumed by the course proposers. This section seeks to establish the reasons why the course was developed; the nature and organisation of the course content; the explicit, aims and objectives of the course; the proposed teaching and learning strategies; and finally the methods of assessment and/or examination of course outcomes.

Section 4 This section is evaluative in the sense that the course materials are re-examined in the light of the previous sections to establish the degree to which the course as written exhibits internal consistency between its objectives, content and methods.

Section 5 A summary of the previous sections which seeks to identify the key questions for the performance and context evaluation.

The full scheme is summarised in Table 1.

Table 1. Summary of the scheme for the analysis and intrinsic evaluation of the curriculum materials

Section 1 Description of the course materials

1.1 Materials for the teacher
1.2 Materials for the pupils
1.3 Other support materials

Section 2. Antecedent conditions

2.1 Pupil age and ability range
2.2 Previous knowledge and experience of the pupils
2.3 Organisation of the teaching group
2.4 Teacher capabilities and requirements
2.5 Curricular implications
2.6 Financial and resource implications

Section 3. Rationale and strategy

3.1 General rationale
3.2 Aims and objectives of the course
3.3 Teaching and learning mode
3.4 Course content
3.5 Teaching and learning strategy
3.6 Internal and external assessment

Section 4. Intrinsic evaluation

4.1 The organisation of the course content
4.2 The relationship between content, techniques, principles and processes
4.3 Teaching modes and interactions
4.4 Homework assignments
4.5 Outcomes of alterrative teaching strategies

Section 5. Summary and implications

5.1 General summary of Section 1 - 4
5.2 Implications for the performance evaluation
5.3 Implications for the context evaluation

In Section 4 use can be made of a quantitative technique for curriculum analysis developed by Easley, Jenkins and Ashenfelter (42). This technique involves the classification of the course content into discrete units, assignable units, which can be set against suitable

42. Easley, J.A., Jenkins, E.S., Ashenfelter, J.W. 'A scheme for the analysis of elementary science materials', EPIE Forum Vol. 1, no. 3 (November 1967) pp. 16 - 21.

descriptors of teaching mode, interaction, or method of presentation. The resultant classification can then be shown in the form of a profile and different parts, or stages, of the course compared. An example of a typical profile is given in Appendix 2 but it should be noted that a classification of this type can only result in a general picture of the pattern of a course or units as the classification procedures used introduce two forms of distortion. Firstly, the translation of statements often presented descriptively in curriculum materials into more specific assignable units introduces a range of approximations and potential inaccuracies. Secondly, the classificatory process itself imposes certain constraints, relies heavily on the skill and judgement of the analyst, and raises questions of validity. Nevertheless, providing these problems are acknowledged, and the conclusions drawn are treated with some caution, the analysis does give valuable insight into the internal structure and balance of the materials.

6.2 The performance evaluation

In the review of models for performance evaluation discussed earlier the general position adopted was to favour a psychometric rather than anthropological, interactional or economic model providing the limitations of the approach are noted and contextual factors are considered in their own right.

In summarising the tactical decisions to be considered in the design of a performance evaluation programme it is suggested that the main objectives of the overall study should be regarded as more important than the development and evaluation of a highly sophisticated statistical methodology, providing the limitations of the techniques

used are made explicit. The performance evaluation is part of a wider study in which the tests of achievement and attitude

> .....enjoy no privileged status within the study. Test scores cannot be considered in isolation; they form merely one section of the data profile. Interest lies not so much in relating different test scores, but in accounting for them using the study's findings as a whole(43).

An intrinsic evaluation utilising the scheme described above will identify the main objectives of the curriculum materials and a range of additional outcomes that arise, or are likely to arise, as a result of course implementation. These objectives and outcomes are the goals that will be evaluated in performance terms, a procedure that involves intranslating statements of intent and expectation into specific performance objectives; selecting or developing appropriate measures of the level of achievement of those objectives; and analysing and presenting the data obtained from observation and measurement in a manner that is appropriate to both the goals being evaluated and the conditions under which the observations were made. In operational terms the transition from the list of performance objectives derived from the intrinsic evaluation to the performance tests to be used in the field involves

a) establishing a suitable classification of objectives;

b) selecting, or writing, test items specific to those objectives;

c) pre-testing and evaluating test items, and the compilation of the final tests used in the study.

The classification of objectives can be based on Bloom's taxonomy objectives modified to reflect the range of objectives being evaluated (44)

---

43. Parlett, M., Hamilton, D. Evaluation as Illumination; a new approach to the study of innovatory programs, (Edinburgh, Centre for Research in the Educational Sciences, 1972) p.22.

44. Bloom, B.S. (Ed.) Taxonomy of Educational Objectives: Book 1 The Cognitive Domain (London, Longmans Green, 1956); and Bloom, B.S. Hastings, J.T., Madaus, G.F. Handbook on Formative and Summative Evaluation of Student Learning (New York, McGraw-Hill, 1971). In the latter book objectives are keyed to the varying needs of different subject areas.

or to an appropriate classification derived from the intrinsic evaluation. By linking test items to specific objectives the cognitive tests utilised in the study are criterion-referenced and assumptions regarding reliability and validity associated with the methodology of norm-referenced tests are questioned (45). Nevertheless test items selected or written should be effectively pre-tested and the results systematically analysed as a first step towards the production of valid evaluation instruments. Where objective tests are used techniques of item analysis are particularly useful in developing tests with an appropriate range of difficulty and the reader is referred to Ebel's valuable handbook for guidance (46). Following the work of Nuttall and Willmott (47) similar considerations can also be applied when open-ended and essay type questions are used.

The utilisation of a taxonomy as a guideline for test development also raises important questions regarding the degree to which different test items measure specific, rather than general abilities, and the hierarchical relationship of those abilities. Whilst note should be taken of several attempts to validate Bloom's taxonomy (48) and to write process specific test items (49), it is not within the terms of

45. Astin, A.V. 'Criterion-centred research', Educational and Psychological Measurement, 24 (1964) pp. 807 - 822.

46. Ebel, R.L. Measuring Educational Achievement (New Jersey, Prentice Hall, 1965).

47. Nuttall, D.L., Willmott, A.S., British Examinations: Techniques of Analysis (Slough, NFER, 1972)

48. See Kropp, R.P., Stoker, H.W. The construction and validation of of tests of the cognitive processes as described in the Taxonomy of Educational Objectives (Florida, Institution of Human Learning and Department of Educational Research and Testing, Florida State University, 1966); and Smith, R.B. 'An empirical examination of the assumptions underlying the Taxonomy of Educational Objectives: Cognitive Domain', Journal of Educational Measurement, Vol. 5, No. 2 (Summer 1968) pp. 125-127.

49. Lewis, D.G. 'Ability in Science at Ordinary Level of the General Certificate of Education' British Journal of Educational Psychology, 35 (1967) pp. 361 - 370.

reference of this paper to do more than acknowledge the possible nature of the assumptions that might be made in the development of tests for a performance evaluation.

The complete development and validation of a performance test of a range of cognitive and affective objectives would involve the factor analysis of the results in order to determine the inter-dependence of categories (50), and techniques such as McQuitty's hierarchical syndrome analysis (51) to establish the order in which the categories were related. In many studies pragmatic decisions relating to time and general resources may well prevent this degree of development and the inter-dependence of categories will have to be assumed.

A further set of tactical decisions that create problems with respect to the analysis and interpretation of the data derived from a performance evaluation are encountered when one considers the selection of samples of pupils for testing. In broad terms the evaluator has to decide whether the organisation of his performance evaluation can meet the tight requirements of a true experimental design or, as an alternative to conceptualise it as an observational study (52). In the former case an appropriate sampling procedure must be adopted which meets the statistical requirements suggested by authorities such as Moser (53). In general this is likely to mean

50. Factor analysis would establish the extent to which each sub-test was independent, i.e. measuring a distinctive quality, but would not in itself validate the notion of a hierarchical, or taxonomic, relationship between factors.

51. McQuitty, L.L. 'Improved hierarchical syndrome analysis of discrete and continuous data' Educational and Psychological Measurement, 26 (1966), pp. 577 - 582.

52. This whole issue is comprehensively reviewed in Kerlinger, F.N. Foundations of Behavioural Research: Educational and Psychological Inquiry (London, Holt, Rinehart and Winston, 1969)

53. Moser, C.A. Survey Methods in Social Investigation (London, Heinemann, 1958).

the evaluator has to work with schools and teachers with whom he has had little previous contact and the level of their active participation in the study is likely to be low. The alternative is to select schools on the basis of their known characteristics and potential support and involvement, particularly where the demands of the contextual evaluation are likely to be heavy, and accept certain constraints in terms of the generalisability of findings. Thus an observational design utilising the testing of complete teaching groups as found in the schools participating in the study is easier to set up than a design which requires either major modifications to teaching groups to create viable control groups, or the testing of sub-sets of pupils derived from a range of teaching groups. Whilst the observational nature of the data does not prevent the use of standard statistical methods of analysis, the frequent failure to obtain equal, or proportionate, numbers of pupils in each test category militates against the use of standard analysis of variance techniques. Snedecor, Winer, and Edwards (54) have, amongst others, suggested viable methods for overcoming many of the problems associated with such designs, all of which are intended to provide better control or estimates of error than, for example, the often used reliance on a series of independent t-tests of group means.

6.3 The context evaluation

The context evaluation is essentially concerned with estimating the effect the implementation of a specific set of course proposals has had on the curriculum in general, together with the extent to

54. Snedecor, G.W. Statistical Methods (Arris, Iowa State College, 1956); Winer, B.J. Statistical Principles in Psychological Design (New York, McGraw-Hill, 1962)- and Edwards, A.L. Experimental Design in Psychological Research (London, Holt, Rinehart and Winston, 1968).

which the course proposals have themselves been modified by the school. The essentially reflexive nature of the interaction between curriculum development and curriculum adoption and implementation provides a fertile area of inquiry in any comprehensive evaluation of an innovation. Amongst the wide range of areas that might be investigated the following represent the most immediately important and are presented as examples in the development of a more systematic approach to context evaluation.

6.31 Effects on teaching and learning styles and methods

New curricular proposals often suggest major changes in teaching method and learning styles as well as changes in subject content. For example the Nuffield Science Teaching Project 'O' Level proposals adopt a teaching method based on a high level of pupil activity in practical work, whilst the Schools' Council Humanities Project places a premium on discussion methods. Evaluating such innovations requires the detailed observation and analysis of the interactions between teachers and pupils, pupils and pupils, and pupils and resources in order to ascertain the extent to which traditional patterns of teacher-pupil interaction have been modified. This is particularly important if one conceptualises a curriculum innovation in terms other than simply a change in either the nature or organisation of subject content.

6.32 Changes in the aims, objectives and attitudes of teaching staff

Whilst the formal measurement of attitude changes in pupils is legitimately considered as part of a performance evaluation, a comparable study of teacher attitudes provides the context for interpreting any results obtained. Central to such a study is the evaluation of the extent to which the explicit aims and objectives of a new curriculum have been accepted and internalised by the teaching staff, and the

extent to which other, not necessarily older, aims have been superimposed on the curriculum. Consideration of these questions should again not be limited to content objectives and the gap, if any, between stated pedagogical aims and actual teaching practice should be carefully examined. Techniques involving the rating of aims according to their ascribed importance as used by Kerr (55) in his study of practical work in school science are useful in this sensitive area of inquiry. In many respects this aspect of the proposed evaluative programme is most open to interference between the observer and the observed, and any technique short of total participant observation can do little other than provide the broadest of indications of trends and positions.

6.33 Effects on the organisation of teaching groups

Apart from the direct effects on curriculum change in terms of the patterns of interaction in the classroom, the evaluator should also consider changes that have occurred in the composition of teaching groups. Has, for example, a subject become more selective as a result of curriculum innovation, with fewer 'less-able' pupils opting to study it? Has the innovation resulted in a different pattern of streaming or setting, or alternatively has a programme developed for more able pupils been implemented across a far wider ability range to that intended? Issues such as these which can be examined through a longitudinal survey of school records are clearly important in the overall evaluation of a curriculum and are related to the broader issues of the curriculum discussed below.

55. Kerr, J.F. Practical Work in School Science (Leicester, Leicester University Press, 1963)

6.34 Effects on the organisation of the curriculum

Here the evaluator is concerned with changes in time allocations, the patterning of subject options, the range and choice of options and potential incompatabilities in approach between one subject or another. For example, the adoption of a set of curriculum proposals based on subject specialisation may create serious problems in a school moving towards integration in other curricular areas and this would be an important 'cost' to evaluate in overall terms. Similarly the acceptance of heuristic teaching methods in one subject may create effects in others, or conversely lead to the failure to properly implement a heuristic approach. Clearly the extent to which any one evaluation study can involve an in-depth consideration of these wider contextual factors is limited, but an awareness that curriculum change in one area resonates in others is important.

6.35 Effects on examination and test procedures and methods

It is important to consider changes in examination procedures that arise, or should arise, from curriculum innovation. Thus the analysis of examination and test papers provides a useful indicator of the extent to which real change has taken place. This suggests that test and examination questions should be classified on the same basis as the objectives derived from the intrinsic evaluation and results of tests and examinations compared with the results of the performance evaluation. In many respects this is one area where major inconsistencies between intentsions and outcomes are likely to arise, especially when innovations involve more than a content shift. In addition to considering the nature of tests and examinations used the evaluator might also consider factors such as the frequency and distribution of testing, and the cost and production difficulties created by some assessment methods.

6.36 Effects on the deployment, utilisation and training of staff

An essential contextual consideration in any evaluation should be the examination of effects of curricular change on the deployment of both teaching and ancillary staff. This may involve changes in staffing patterns and flexibility, important changes in the level of in-service training within a school; and possible restrictions in the deployment of teachers not formally involved in the innovation. Clearly these considerations apply to ancillary staff such as media resource officers, clerical assistants and laboratory and workshop technicians. In an extreme case the evaluator may establish that one of the effects of efficiently staffing an innovation may have been to denude other subject areas of essential support. Such a finding would play an important part in cost-benefit or cost-effectiveness terms in the overall evaluation of the innovation.

6.37 Effects on the allocation and utilisation of resources

Whilst the allocation and deployment of teaching and ancillary staff is a major aspect of resource utilisation generally, it is helpful to regard factors such as equipment, books, materials and teaching and learning spaces under a separate heading although the questions and issues discussed in the paragraph above equally apply. Again the findings of an analysis of resource provision are best considered in cost-benefit or cost-effectiveness terms.

6.38 Summary & methodological implications

Whilst the compilation of data under the range of headings suggested above presents no conceptual or methodological problems additional to those discussed in the sections on intrinsic and performance

evaluation, it is clear that the analysis and evaluation of the data does. By covering a wide range of factors the volume of data tends to be large and this is compounded if a large number of schools are involved in any one study. In addition comparative data on factors such as the deployment of staff and resources derived from a range of often very different schools is difficult to evaluate as norms are not available. The evaluator is therefore forced back to an analysis and presentation that can only be crudely quantitative and comparative, and which will require him to make his own value assumptions fully explicit. On many issues probably the best tactic is to attempt to map changes that have occurred over time, preferably by incorporating a time-span that goes back beyond the date of the innovation being studied. Thus a 'before and after' analysis of time allocations, curriculum patterns, staff deployments, and cash and materials allocations allows the evaluator to paint a picture of the effects of an innovation in terms of decisions made by teachers and administrators. The nature of these decisions in itself illuminates aspects of the value positions, aims and objectives of the staff concerned and this view can be set against any formal analysis of stated attitudes and aims. A direct longitudinal survey of staff attitudes and any changes resulting from the implementation of innovations is rarely possible as time and staff mobility present major problems.

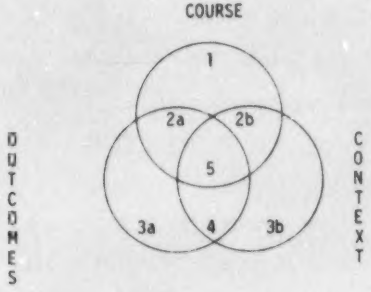
One may therefore summarise by suggesting that the context evaluation aims to present an illuminative gloss on the 'harder' data obtained from a performance evaluation. It involves both clarifying the context from which the performance data was obtained and creating a context within which it can be more fully evaluated. Thus the 'hard'

fact that on average the attainment quotients of fourth form boys following Course A are ten points higher than those following Course B is illuminated by knowing something of the context within which those boys are taught and work, and can be more critically evaluated in the context of the concomitant changes that might have been made in order to achieve those standards.

7.0 SUMMARY AND CONCLUSIONS

This paper has sought to clarify the aims and methodology of summative curriculum evaluation by outlining a framework within which a series of strategic and tactical decisions may be taken. The model described schematically in Figure 1 consists of three stages - intrinsic, performance and context evaluations - each with a distinct point of focus.

Figure 1. Schematic representation of the proposed summative evaluation process



	FOCUS	PROCESS	ACTIVITIES
1.	COURSE	INTRINSIC EVALUATION	Description; analysis of rationale and strategy; analysis of aims and objectives; intrinsic evaluation.
2a	COURSE/ OUTCOMES	HYPOTHESIS FORMATION	Specification of the objectives of the performance evaluation.
2b	COURSE/ CONTEXT	HYPOTHESIS FORMATION	Specification of the objectives of the context evaluation.
3a	OUTCOMES	PERFORMANCE EVALUATION (Hypothesis testing)	Classification of objectives; selection production and pilot testing of tests, scales, the questionnaires; selection of samples; testing and data collection; analysis of data; interim findings.
3b	CONTEXT	CONTEXT EVALUATION (Hypothesis testing)	Classification of objectives; selection, production and pilot testing of questionnaires- data collection; analysis of data; interim findings.
4.	OUTCOMES/ CONTEXT	ANALYSIS OF INTERACTION (Evaluation)	Examination of the interaction between performance and context and re-evaluation of both sets of findings.
5.	COURSE/ CONTEXT	SUMMATIVE EVALUATION	Re-examination of results in terms of the original intrinsic evaluation. Final report and recommendations.

In the intrinsic evaluation the central concern is the course, or curriculum, which is described and analysed in order to determine intended and unintended outcomes in terms of pupil performance and contextual effects. This process essentially forms the hypotheses which are tested in the two subsequent stages of the evaluation. The focus for the performance evaluation is the analysis of pupil attainments and attitudes on the range of objectives derived from the intrinsic evaluation. Whilst a wide range of possible approaches to this stage have been discussed earlier in the paper, it should be stressed that many problems associated with performance evaluation techniques have not been dealt with exhaustively. The full treatment of, for example, the statistical methodology for non-experimental or quasi-experimental designs is outside the terms of reference of an introductory paper, but the writer hopes that by setting proposals within a substantive survey of the literature suitable guidance has been made available. A similar stance has been adopted with respect to the discussion of the context evaluation, where the focus is the educational environment within which the course has been implemented and the pupils are taught. Again specific hypotheses generated from the intrinsic evaluation are tested.
The scheme is completed by a detailed consideration of the interaction between performance and context, and the interpretation of findings in terms of the original intrinsic evaluation. It is suggested that findings would take the form of descriptive and quantitative statements about levels of performance linked to suggestions for either improving course structures, curriculum materials and resources, or, where appropriate, modifying implementation strategies. The final report would ideally be framed to assist decision makers in the on-going process of curriculum improvement and where this is achieved and decisions are made the formative dimension of an overtly summative strategy becomes explicit. Ultimately this is the acid test of any curriculum evaluation exercise.

APPENDIX 1

GENERAL BIBLIOGRAPHY

A. Evaluation Methodology

Beatty, W.H. (Ed.) — Improving Educational Assessment and An Inventory of Measures of Affective Behaviour (Washington, Association for Supervision and Curriculum Development, N.E.A. 1969).

Bloom, B.S., Hastings, J.T., Madaus, G.F. — Handbook of Formative and Summative Evaluation of Student Learning, (New York, McGraw-Hill, 1971)

Kraft, R.H.P. et al — Four Evaluation Examples: Anthropological, Economic, Narrative and Portrayal, (Chicago, Rand McNally, 1974).

Grobman, H. — Evaluation Activities of Curriculum Projects, (Chicago, Rand McNally, 1968).

Parlett, M., Hamilton, D. — Evaluation as Illumination: a new approach to the study of innovatory programs (University of Edinburgh, Centre for Research in the Educational Sciences, Occasional Paper 9, 1972, mimeo).

Stufflebeam, D.L. et al — Educational Evaluation and Decision Making (Itasca, Peacock, 1971)

Tyler, R.W. (Ed.) — Educational Evaluation: New Roles, New Means (Chicago, University of Chicago Press, 1969)

Tyler, R.W., Gagne, R.M., Scriven, M. — Perspectives of Curriculum Evaluation (Chicago, Rand McNally, 1967)

Wittrock, M.C., Wiley, D.E. — The Evaluation of Instruction: Issues and Problems (New York, Holt, Rinehart & Winston, 1970)

B. Evaluation Studies

Alexander, D.J. — Nuffield Secondary Science: An Evaluation (London, Macmillan Education, 1974)

Crossland, R.W., Moore, S.F.D. — Environmental Studies Project 5-13: An Evaluation (London, Macmillan Education, 1974)

Lindvall, C.M., Cox, R.C. — The IPI Evaluation Program (Chicago, Rand McNally, 1970)

Schools' Council — Evaluation in Curriculum Development: Twelve Case Studies (London, Macmillan Education, 1973)

Shipman, M., Bolam, D., Jenkins, D.R. — Inside a Curriculum Project (London, Methuen, 1974)

C. General Methodology

C1. Attitude Measurement

Edwards, A.L. — Techniques of Attitude Scale Construction (New York, Appleton, 1956)

Fishbein, M. — Readings in Attitude Theory and Measurement (New York, Wiley, 1967)

Oppenheim, A.N. — Questionnaire Design and Attitude Measurement (London, Heinemann, 1966)

Triandis, H.C. — Attitude and Attitude Change (London, Wiley, 1971)

C2. Classroom Interaction

Bellack, A.A. et al — The Language of the Classroom (New York, Teachers' College Press, 1966)

Eggleston, J.F., Galton, M., Jones, M. — Contexts of Science Learning (Leicester, University of Leicester School of Education, 1973, mimeo)

Flanders, N.A. — Analysing Teaching Behaviour (New York, Addison-Wesley, 1970)

Gallagher, J.J., Nuthall, G.A., Rosenshine, B. — Classroom Observation (Chicago, Rand McNally, 1970)

C3. Curriculum Analysis

Eraut, M.R., Goad, L.H., Smith, G. — The Analysis of Curriculum Materials (Sussex, University of Sussex Education Area, Occasional Paper No. 2, 1975, mimeo)

Hutchins, C.C. (Ed.) — Science - A Process Approach: Programme Report (Berkeley, Far West Laboratory for Educational Research and Development, 1970)

C4. Experimental Design and Sampling Techniques

Campbell, D.T., Stanley, J.C. — Experimental and Quasi-Experimental Designs for Research on Teaching (Chicago, Rand McNally, 1965)

Edwards, A.L. — Experimental Design in Psychological Research (New York, Holt, Rinehart and Winston, 1968)

Kerlinger, F.N. — Foundations of Behavioral Research (New York, Holt, Rinehart and Winston, 1964)

Lewis, D.F. — Experimental Design in Education (London, University of London Press, 1968)

Lindquist, E.F. — Design and Analysis of Experiments in Psychology and Education (Boston, Houghton Mifflin, 1956)

Moser, C.A. — Survey Methods in Social Investigation (London, Heinemann, 1958)

C5. Statistical Methods

Edwards, A.L. — Statistical Analysis (New York, Holt Rinehart and Winston, 1958)

Furguson, G. — Statistical Analysis in Psychology and Education (New York, McGraw-Hill, 1966)

Hays, W.L. — Statistics for Psychologists (New York, Holt, Rinehart and Winston, 1964)

Siegel, S. — Non-parametric Statistics for the Behavioral Sciences (New York, McGraw-Hill, 1966)

Winer, B.J. — Statistical Principles in Psychological Design (New York, McGraw-Hill, 1962)

C6. Test Construction and Methodology

Bloom, B.S. (Ed.) — Taxonomy of Educational Objectives: Book 1 The Cognitive Domain (London, Longmans, 1964)

Cox, R.C. Jordan, J.M. — Validation and Uses of the Taxonomy of Educational Objectives Cognitive Domain (Pittsburgh, University of Pittsburgh School of Education, 1965)

Ebel, R.L. — Measuring Educational Achievement (New Jersey, Prentice-Hall, 1965)

Gronlund, N.E. — Constructing Achievement Tests (Englewood Cliffs, Prentice-Hall, 1968)

Krathwohl, D.R. (Ed.) — Taxonomy of Educational Objectives: Book 2 The Affective Domain (London, Longmans, 1964)

Kropp, R.P., Stoker, H.W. — The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives (Florida, Institution of Human Learning and Department of Educational Research and Testing, Florida State University, 1966)

Nuttall, D.L., Wilmott, A.S. — British Examinations: Techniques of Analysis (Slough, NFER, 1972)

C7. General

Becker, H.S. — Problems of Inference and Proof in Participant Observation (London, Allen Lane, The Penguin Press, 1971)

Gage, N.L. (Ed.) — Handbook of Research on Teaching Chicago, Rand McNally, 1963)

Lindquist, E.F. (Ed.) — Educational Measurement (Washington D.C., American Council on Education, 1951)

Travers, R.M.W. (Ed.) — The Second Handbook of Research on Teaching (Chicago, Rand McNally, 1973)

APPENDIX 2 A Sample Curriculum Profile

The Nuffield Science Teaching Project Ordinary Level Chemistry Proposals are organised into units each of which can be classified as being primarily concerned with

A. obtaining new materials from those already available;
B. acquiring experimental and manipulative skills;
C. identifying patterns in the behaviour of substances;
D. understanding and explaining patterns in the behaviour of substances;
E. associating energy changes with material changes;
F. examining the relevance of chemistry to man's social and needs (1)

These six categories represent the headings under which each of the topics/activities presented in the course handbook The Sample Scheme (2) can be assigned (These are designated in Table 2 by the respective code numbers as presented in the Sample Scheme i.e. each topic represents an 'assignable unit'). In operating the Easley, Jenkins and Ashenfelter schemethe evaluator assigns each topic/unit to an appropriate category, as shown in Table 2. The results of the analysis can either be shown as percentages within each major course area, Table 3 or graphically as in Figure 2. (In both cases Stages 1A and II represent different vertical divisions of the course.) Differences in emphasis between categories and divisions of the course, or between courses in a comparative evaluation, are then readily made.

---

1. See the Nuffield Science Teaching Project 'O' Level Chemistry: Introduction and Guide (London, Longmans/Penguin, 1966)

2. The Nuffield Science Teaching Project 'O' Level Chemistry: The Sample Scheme Stages I and II, The Basic Course (London, Longmans/Penguin, 1966).

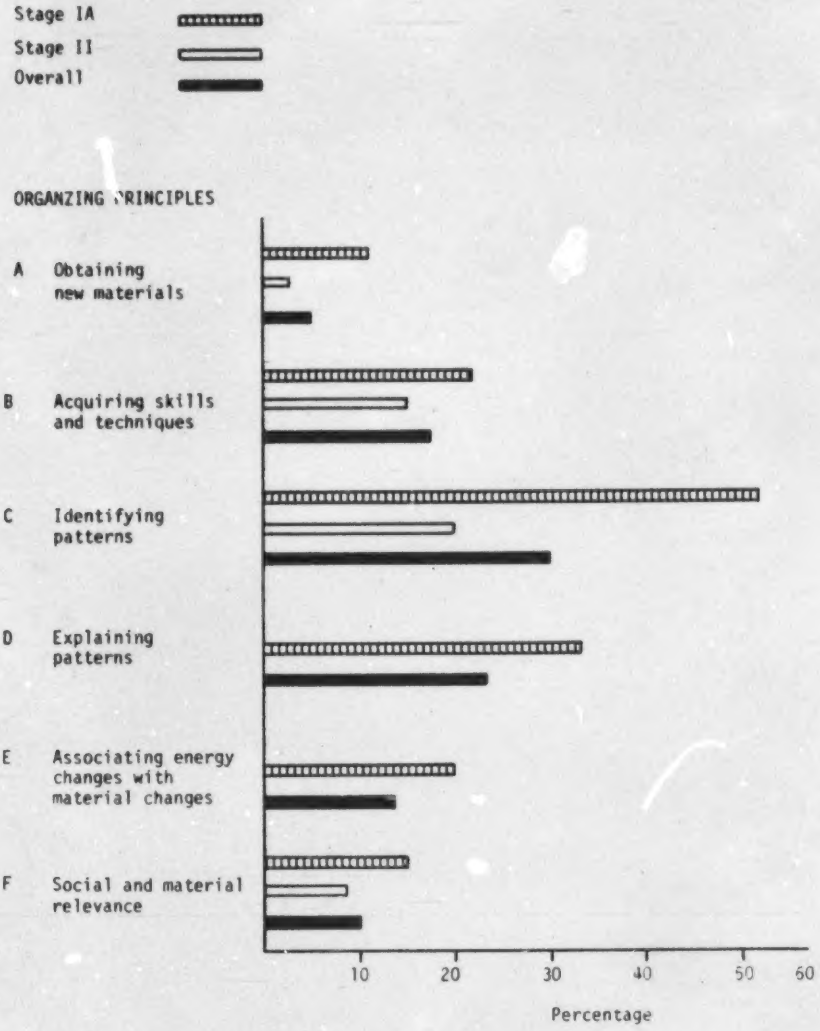
Table 2 Classification of assignable units in terms of the main organising principles of the course content

Category	Stage IA units	Stage II units	Totals		
			I	II	All
A Obtaining new materials	1.1;3.1;10.2;	12.1;22.3;	3	2	5
B Acquiring skills and techniques	1.2;2.1;3.2; 8.1;8.4;10.3;	11.7;11.8;12.2; 17.1;17.2;17.3; 18.1;20.4;23.8; 23.9;	6	10	16
C Identifying Patterns	2.2;2.3;2.4; 2.5;4.1;4.2; 4.3;5.1;5.2; 5.3;6.1;8.2; 8.3;10.1;	13.2;13.3;13.4; 13.5;16.1;16.3; 16.5;16.7;20.1; 21.1;21.4;24.1; 24.2;	14	13	27
D Explaining Patterns		11.1;11.5;11.6; 13.1;14.1;14.2; 14.3;14.4;15.1; 16.2;16.4;16.6; 16.8;17.4;18.2; 18.5;18.6;19.1; 19.2;20.2;20.3;	0	21	21
E Associating energy changes with material changes		15.2;15.3;15.4; 15.5;18.3;18.4; 22.4;23.2;23.3; 23.4;23.5;23.6; 23.7;	0	13	13
F Social and material relevance	1.4;1.5;9.1; 9.2;	21.2;21.3;22.1; 22.2;23.1;24.3;	4	6	10
Units not classified	1.3;2.6;3.3; 3.4;6.2;6.3; 7.1;7.2;7.3; 7.4;7.5;	11.2;11.3;11.4; 22.5;	27 11	65 4	92 15
TOTALS			38	69	107

Table 3. Units in each course content category expressed as a percentage of the total units classified in each stage of the course.

Category	% Stage I	% Stage II	% Total
A	11	3	5.5
B	22	15	17.5
C	52	20	30
D	0	33	23
E	0	20	14
F	15	9	10
	100	100	100

Figure 2. Course profile expressed in terms of the main organizing principles of the course content



The Author

Richard West trained as a teacher at the College of St. Mark and St. John, Chelsea and from 1958-1970 taught science at Walworth Comprehensive School, London where for six years he was Head of the Science Department. In 1970 he joined the permanent staff of the Centre for Educational Technology, University of Sussex with a responsibility for developing regional support services for teachers. From 1971-1973 he coordinated and led curriculum development programmes for a new regional B.Ed. degree for serving teachers, and for two years was course director of the M.A. in Curriculum Development and Educational Technology. His personal research interests in science education were developed concurrently and he was awarded a D.Phil. in 1974 for his evaluation of the Nuffield Science Teaching Project Ordinary Level Chemistry proposals. He is currently directing a three year research study of Environmental Education in primary and secondary schools sponsored by the Leverhulme Trust, and is Chairman of the Education Research Committee of the Association for Science Education.